

On sampling theorem with sparse decimated samples: exploring branching spectrum degeneracy

Nikolai Dokuchaev

Submitted: May 2, 2016. Revised: February 28, 2017

Abstract

The paper investigates possibility of recovery of sequences from their decimated subsequences. It is shown that this recoverability is associated with certain spectrum degeneracy of a new kind, and that a sequences of a general kind can be approximated by sequences featuring this degeneracy. This is applied to sparse sampling of continuous time band-limited functions. The paper shows that these functions allow an arbitrarily close approximation by functions that can be recovered from sparse equidistant samples with sampling distance larger than the distance defined by the critical Nyquist rate for the underlying function. This allows to bypass, in a certain sense, the restriction on the sampling rate defined by the Nyquist rate.

Keywords: sampling, bandlimitness, missing values, one-sided sequences, sparse samples.

MSC 2010 classification : 94A20, 94A12, 93E10

1 Introduction

The paper investigates possibility of recovery of sequences from their decimated subsequences. It appears there this recoverability is associated with certain spectrum degeneracy of a new kind, and that a sequences of a general kind can be approximated by sequences featuring this degeneracy. This opens some opportunities for sparse sampling of continuous time band-limited functions. As is known, the sampling rate required to recovery of function is defined by the classical Sampling Theorem also known as Nyquist-Shannon theorem, Nyquist-Shannon-Kotelnikov theorem, Whittaker-Shannon-Kotelnikov theorem, Whittaker-Nyquist-Kotelnikov-Shannon theorem, which is one of the most basic results in the theory of signal processing and information science; the result was obtained independently by four authors [25, 13, 12, 19]. The theorem states that a band-limited function can be uniquely recovered

The author is with Department of Mathematics and Statistics, Curtin University, GPO Box U1987, Perth, Western Australia, 6845 (email N.Dokuchaev@curtin.edu.au).

without error from a infinite two-sided equidistant sampling sequence taken with sufficient frequency: the sampling rate must be at least twice the maximum frequency present in the signal (the critical Nyquist rate). Respectively, more frequent sampling is required for a larger is the spectrum domain. If the sampling rate is preselected, then it is impossible to approximate a function of a general type by a band-limited function that is uniquely defined by its sample with this particular rate. This principle defines the choice of the sampling rate in almost all signal processing protocols; as far as we know, nobody even considered yet a possibility to bypass it. Numerous extensions of this theorem were obtained, including the case of nonuniform sampling and restoration of the signal with mixed samples; see some recent literature review in [1, 11, 22, 24, 26]. There were works studying possibilities to reduce the set of sampling points required for restoration of the underlying functions. In particular, it was found that a bandlimited function can be recovered without error from an oversampling sample sequence if a finite number of sample values is unknown, or if an one-sided half of any oversampling sample sequence is unknown [23]. It was found [10] that the function can be recovered without error with a missing equidistant infinite subsequence consistent of n th member of the original sample sequence, i.e. that each n th member is redundant, under some additional constraints on the oversampling parameter. The constraints are such that the oversampling parameter is increasing if $n \geq 2$ is decreasing. There is also an approach based on the so-called Landau's phenomenon [14, 15]; see [2, 14, 15, 16, 20, 21] and a recent literature review in [17]. This approach allows arbitrarily sparse discrete uniqueness sets in the time domain for a fixed spectrum range. Equivalently, given a minimal distance between sampling points, this approach allow to find an uniqueness set for classes of functions with an arbitrarily large measure of the spectrum range. The corresponding sampling times depend on the spectrum range; they are not equidistant and their calculation is can be complicated. These results are focused on the uniqueness problem rather than on possibility of numerical implementation. As was emphasized in [15], the corresponding do not lead to a stable data recovery; moreover, any sampling below the Nyquist rate cannot be stable in this sense.

The present paper readdresses the problem of reducing the sample required for reconstruction of the underlying continuous time band-limited function. It appears that analysis of the degeneracy of the classical frequency characteristic such as Z-transform and DFT is insufficient for overcoming these restrictions.

To solve the problem, we introduce f some special type of spectrum degeneracy for sequences such that, for a given integer $m > 0$, the processes $\tilde{x} \in \ell_2$, featuring this degeneracy, must have the following properties:

- (i) \tilde{x} can be recovered from a subsample $\tilde{x}(km)$;
- (ii) these processes are everywhere dense in ℓ_2 , i.e. they can approximate any $x \in \ell_2$.

Let us describe briefly introduced below classes of processes with these properties. For a process $x \in \ell_2$, we consider an auxiliary "branching" process, or a set of processes $\{\hat{x}_d\}_{d=-m+1}^{m-1} \in \ell_2$ approximating alterations of the original process. It appears that certain conditions of simultaneous degeneracy of the spectrums for all \hat{x}_d , ensures that it is possible to compute a new \tilde{x} with desired properties (i)-(ii). The set of corresponding sequences \tilde{x} can be considered as a set of sequences featuring spectrum degeneracy of a new kind (Theorem 1–3 below). Moreover, the procedure of recovery of any finite part of the sample $\tilde{x}(k)$ from the subsample $\tilde{x}(km)$ feature some robustness with respect to noise contamination and data truncation, i.e. is a well-posed problem in a certain sense (Theorems 4). Continuous time band-limited functions defined by their samples \tilde{x} from this class also feature spectrum degeneracy of a new kind. This implies that an arbitrarily small adjustment of a band-limited underlying continuous time function convert it to a function that is uniquely defined by its equidistant sample with the sampling distance m times larger than the distance defined by the critical Nyquist rate (Theorems 6 and 7). This allows to bypass, in a certain sense, the restriction on the sampling rate described by the critical Nyquist rate. In addition, application of this approach to more general non-bandlimited functions allows to approximate them by functions that are recoverable from equidistant samples with arbitrarily large distance between sampling points (Corollary 1). This may lead to new approaches to the aliasing problem arising in time discretization of non-bandlimited continuous time functions.

The proof is based on enchanted predictability of sampling sequences satisfying augmented degeneracy conditions. We use a linear predictor representing a modification of the predictor from [3] (see the proof of Theorem 1).

Some definitions

We denote by $L_2(D)$ the usual Hilbert space of complex valued square integrable functions $x : D \rightarrow \mathbf{C}$, where D is a domain.

For $f \in L_2(\mathbf{R})$, we denote by $F = \mathcal{F}f$ the function defined on $i\mathbf{R}$ as the Fourier transform of f ;

$$F(i\omega) = (\mathcal{F}f)(i\omega) = \int_{-\infty}^{\infty} e^{-i\omega t} f(t) dt, \quad \omega \in \mathbf{R}.$$

Here $i = \sqrt{-1}$ is the imaginary unit. For $f \in L_2(\mathbf{R})$, the Fourier transform F is defined as an element of $L_2(i\mathbf{R})$, i.e. $F(i\cdot) \in L_2(\mathbf{R})$.

For $\Omega > 0$, let $L_2^{BL, \Omega}(\mathbf{R})$ be the subset of $L_2(\mathbf{R})$ consisting of functions f such that $f(t) = (\mathcal{F}^{-1}F)(t)$, where $F(i\cdot) \in L_2(i\mathbf{R})$ and $F(i\omega) = 0$ for $|\omega| > \Omega$.

We denote by \mathbb{Z} the set of all integers. Let $\mathbb{Z}^+ = \{k \in \mathbb{Z} : k \geq 0\}$ and let $\mathbb{Z}^- = \{k \in \mathbb{Z} : k \leq 0\}$.

For a set $G \subset \mathbb{Z}$ and $r \in [1, \infty]$, we denote by $\ell_r(G)$ a Banach space of complex valued sequences $\{x(t)\}_{t \in G}$ such that $\|x\|_{\ell_r(G)} \triangleq (\sum_{t \in G} |x(t)|^r)^{1/r} < +\infty$ for $r \in [1, +\infty)$, and

$\|x\|_{r(G)} \triangleq \sup_{t \in G} |x(t)| < +\infty$ for $r = \infty$. We denote $\ell_r = \ell_r(\mathbb{Z})$.

Let $D^c \triangleq \{z \in \mathbf{C} : |z| > 1\}$, and let $\mathbb{T} = \{z \in \mathbf{C} : |z| = 1\}$.

For $x \in \ell_2$, we denote by $X = \mathcal{Z}x$ the Z-transform

$$X(z) = \sum_{k=-\infty}^{\infty} x(k)z^{-k}, \quad z \in \mathbf{C}.$$

Respectively, the inverse Z-transform $x = \mathcal{Z}^{-1}X$ is defined as

$$x(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(e^{i\omega}) e^{i\omega k} d\omega, \quad k = 0, \pm 1, \pm 2, \dots$$

For $x \in \ell_2$, the trace $X|_{\mathbb{T}}$ is defined as an element of $L_2(\mathbb{T})$.

2 The main results

Let us formulate first the classical Nyquist-Shannon-Kotelnikov Theorem: *Let $\Omega > 0$ and $\tau > 0$ be given and $\Omega \leq \pi/\tau$. Let $\{t_k\}_{k \in \mathbb{Z}} \subset \mathbf{R}$ be a sequence such that $t_k = \tau k$, $k \in \mathbb{Z}$. Then a band-limited function $f \in L_2^{BL, \Omega}(\mathbf{R})$ is uniquely defined by the sequence $\{f(t_k)\}_{k \in \mathbb{Z}}$.*

The sampling rate $\tau = \pi/\Omega$ is called the critical Nyquist rate for $f \in L_2^{BL, \Omega}(\mathbf{R})$. If $\tau < \pi/O$, then, for any finite set S or for $S = \mathbb{Z}^\pm$, $f \in L_2^{BL, \Omega}(\mathbf{R})$ is uniquely defined by the values $\{f(t_k)\}_{k \in \mathbb{Z} \setminus S}$; this was established in [9, 23]. We cannot claim the same for some infinite sets of missing values. For example, it may happen that $f \in L_2^{BL, \Omega}(\mathbf{R})$ is not uniquely defined by the values $\{f(t_{mk})\}_{k \in \mathbb{Z}}$ for $m \in \mathbb{Z}^+$, if the sampling rate for this sample $\{f(t_{mk})\}_{k \in \mathbb{Z}}$ is lower than is the so-called critical Nyquist rate implied by the Nyquist-Shannon-Kotelnikov Theorem; see more examples in [10]. We address this problem below.

Up to the end of this paper, we assume that we are given $\Omega > 0$ and $\tau > 0$ such that $\Omega \leq \pi/\tau$. in addition, we are given $m \geq 1$, $m \in \mathbb{Z}$. We will denote $t_k = \tau k$, $k \in \mathbb{Z}$.

A special type of spectrum degeneracy for sequences

We consider first some problems of recovering sequences from their decimated subsequences.

Definition 1 Consider an ordered set $\{x_d\}_{d=-m+1}^{m-1} \in (\ell_2)^{2m-1}$ such that $x_d(k) = x_0(k)$ for $k \leq 0$ and $d > 0$, and that $x_d(k) = x_0(k)$ for $k \geq 0$ and $d < 0$. We say that this is a branching process, and we say that x_0 is its root.

Definition 2 Consider a branching process $\{x_d\}_{d=-m+1}^{m-1} \in (\ell_2)^{2m-1}$. Let $\tilde{x} \in \ell_2$ be defined such that $\tilde{x}(k) = x_d(k+d)$, where d is such that $(k+d)/m \in \mathbb{Z}$ and that $d \in \{0, 1, \dots, m-1\}$ if $k \geq 0$, $d \in \{-m+1, \dots, 0\}$ if $k < 0$. Then \tilde{x} is called the representative branch for this branching process.

For $\delta > 0$ and $n \geq 1$, $n \in \mathbb{Z}$, let

$$\begin{aligned} s_{n,k} &= \frac{2\pi k - \pi}{n}, \quad k = 0, 1, \dots, n-1, \\ J_{\delta,n} &= \{\omega \in (-\pi, \pi] : \min_{k=0,1,\dots,n-1} |\omega - s_{n,k}| \leq \delta\}. \end{aligned} \quad (1)$$

Let $V(\delta, n)$ be the set of all $x \in \ell_2$ such that $X(e^{i\omega}) = 0$ for $\omega \in J_{\delta,n}$, where $X = \mathcal{Z}^{-1}x$.

Definition 3 Let $\delta > 0$ be given. We say that a branching process $\{\hat{x}_d\}_{d=1-m}^{m-1} \in (\ell_2)^{2m-1}$ features branching spectrum degeneracy with parameters (δ, m) if $\hat{x}_d \in V(\delta, \zeta(d))$ for some positive $\zeta(d) \in \mathbb{Z}^+$ such that $\zeta(d)/m \in \mathbb{Z}$, for $d = -m+1, 1, \dots, m-1$. We denote by $\mathcal{U}(\delta, m)$ the set of all branching processes with this feature.

Definition 4 Let $\tilde{x} \in \ell_2$ be a representative branch of a branching process from $\mathcal{U}(\delta, m)$ for some $\delta > 0$. We say that \tilde{x} features branching spectrum degeneracy with parameters (δ, m) . We denote by $U(\delta, m)$ the set of all sequences \tilde{x} with this feature.

Theorem 1 Any $\tilde{x} \in \cup_{\delta>0} U(\delta, m)$ is uniquely defined by the subsequence $\{\tilde{x}(mk)\}_{k \in \mathbb{Z}}$.

Theorem 2 For any branching process $\{x_d\}_{d=-m+1}^{m-1} \in (\ell_2)^{2m-1}$, $m \geq 1$, $m \in \mathbb{Z}$, and any $\varepsilon > 0$, there exists a branching process $\{\hat{x}_d\}_{d=-m+1}^{m-1} \in \cup_{\delta>0} \mathcal{U}(\delta, m)$ such that

$$\|x_d - \hat{x}_d\|_{\ell_2} \leq \varepsilon, \quad d = 0, 1, \dots, m-1. \quad (2)$$

Consider mappings $\mathcal{M}_d : \ell_2 \rightarrow \ell_2$ such that, for $x \in \ell_2$, the sequence $x_d = \mathcal{M}_d x$ is defined as the following:

$$\begin{aligned} \text{(a)} \quad & x_0 = x; \\ \text{(b)} \quad & \text{for } d > 0 : \quad \begin{aligned} x_d(k) &= x(k), \quad k < 0, \\ x_d(k) &= x(0), \quad k = 0, 1, \dots, d, \\ x_d(k) &= x(k-d), \quad k = d+1, d+2, \dots; \end{aligned} \\ \text{(c)} \quad & \text{for } d < 0 : \quad \begin{aligned} x_d(k) &= x(k), \quad k > 0, \\ x_d(k) &= x(0), \quad k = 0, -1, \dots, d, \\ x_d(k) &= x(k+d), \quad k = d-1, d-2, \dots \end{aligned} \end{aligned} \quad (3)$$

Let $x \in \ell_2$, and let a branching process $\{\hat{x}_d\}_{d=-m+1}^{m-1}$ be such that $x_d = \mathcal{M}_d x$. It follows from the definitions that, in this case, the root branch x_0 is also the representative branch. In this case, Theorem 1 implies that the sequence x_0 is uniquely defined by its subsequence $\{x_0(mk)\}_{k \in \mathbb{Z}}$.

Theorem 3 For any $x \in \ell_2$ and any $\tilde{\varepsilon} > 0$, there exists $\tilde{x} \in \cup_{\delta>0} U(\delta, m)$ such that

$$\|x - \tilde{x}\|_{\ell_2} \leq \tilde{\varepsilon}. \quad (4)$$

In particular, this process can be constructed as the following: construct a branching process $\{x_d\}_{d=-m+1}^{m-1}$ such that $x_d = \mathcal{M}_d x$, and, using Theorem 2, find $\{\hat{x}_d\}_{d=-m+1}^{m-1} \in \cup_{\delta>0} \mathcal{U}(\delta, m)$ such that (2) holds with $\varepsilon = \tilde{\varepsilon}/(2m-1)$. Then (4) holds for \tilde{x} selected as the representative branch of $\{\hat{x}_d\}_{d=-m+1}^{m-1}$.

According to Theorem 3, the set $\cup_{\delta>0} U(\delta, m)$ is everywhere dense in ℓ_2 ; this leads to possibility of applications for sequences from ℓ_2 of a general kind.

Furthermore, Theorem 1 represents an uniqueness result in the spirit of [14, 15, 16, 17, 20, 21]. However, the problem of recovery \tilde{x} from its subsequence features some robustness as shown in the following theorem.

Let $B_\delta = \{\xi \in \ell_2 : \|\xi\|_{\ell_2} \leq \delta\}$, $\delta \geq 0$.

Theorem 4 Under the assumptions and notations of Theorem 1, consider a problem of recovery of the set $\{\tilde{x}(k)\}_{k=-M}^M$ from a noise contaminated truncated series of observations $\{\tilde{x}(mk) + \xi(mk)\}_{k=-N}^N$, where $M > 0$ and $N > 0$ are integers, $\xi \in \ell_2$ represents a noise contaminating the observations. This problem is well-posed in the following sense: for any M and any $\tilde{\varepsilon} > 0$, there exists a recovery algorithm and $\delta_0 > 0$ such that for any $\delta \in [0, \delta_0)$ and $\xi \in B_\delta$ there exists an integer $N_0 > 0$ such the recovery error $\max_{n=-M, \dots, M} |\tilde{x}(n) - \tilde{x}(n)|$ is less than $\tilde{\varepsilon}$ for all $N \geq N_0$. Here $\tilde{x}(n)$ is the estimate of $\tilde{x}(n)$ obtained by the corresponding recovery algorithm.

Sparse decimated sampling for continuous time functions

Let consider sparse sampling for continuous time functions.

Theorem 5 For any $\tilde{x} \in U(\delta, m)$, where $\delta > 0$, there exists an unique $\tilde{f} \in L_2^{BL, \pi/\tau}(\mathbf{R})$ such that $\tilde{f}(t_k) = \tilde{x}(k)$. This \tilde{f} is uniquely defined by the sequence $\{\tilde{f}(t_{mk})\}_{k \in \mathbb{Z}}$.

We say that $\tilde{f} \in F(\delta, m)$ is such as described in Theorem 5. Theorem 5 implies that $F(\delta, m)$ can be considered as a class of functions featuring spectrum degeneracy of a new kind.

Theorem 6 Let $\varepsilon > 0$ and $f \in L_2^{BL, \pi/\tau}(\mathbf{R})$ be given. Let $x \in \ell_2$ be defined as $x(k) = f(t_k)$ for $k \in \mathbb{Z}$, and let $\tilde{x} \in \cup_{\delta>0} U(\delta, m)$ and $\tilde{f} \in L_2^{BL, \pi/\tau}(\mathbf{R})$ be defined as described in Theorem 5 such that (4) holds. Then this \tilde{f} is uniquely defined by the values $\{\tilde{f}(t_{mk})\}_{k \in \mathbb{Z}}$, and

$$\|f - \tilde{f}\|_{L_p(\mathbf{R})} \leq C\varepsilon, \quad p = 2, +\infty.$$

where $C > 0$ depends on (Ω, m, τ) only.

Theorem 7 *Under the assumptions and notations of Theorem 6, consider a problem of recovery of the set $\{\tilde{f}(t_n)\}_{n=-M}^M$ from a noise contaminated truncated series of observations $\{\tilde{f}(t_{mk}) + \xi(mk)\}_{k=-N}^N$, where $M > 0$ and $N > 0$ are integers, $\xi \in \ell_2$ represents a noise contaminating the observations. This problem is well-posed in the following sense: for any M and any $\tilde{\varepsilon} > 0$, there exists a recovery algorithm and $\delta_0 > 0$ such that for any $\delta \in [0, \delta_0)$ and $\xi \in B_\delta$ there exists an integer $N_0 > 0$ such the recovery error $\max_{n=-M, \dots, M} |\tilde{f}(t_n) - \tilde{f}_E(t_n)|$ is less than $\tilde{\varepsilon}$ for all $N \geq N_0$. Here $\tilde{f}_E(t_n)$ is the estimate of $\tilde{f}(t_n)$ obtained by the corresponding recovery algorithm.*

Remark 1 *Theorems 5 and 6 represent an uniqueness result in the spirit of [14, 15, 16, 17, 20, 21]. These works considered more general function with unbounded spectrum domain; the uniqueness sets of times in these works are not equidistant and have a complicated structure depending on the spectrum range. We consider band-limited functions f and \tilde{f} only. The novelty is that the uniqueness set $\{t_{mk}\}_{k \leq 0, k \in \mathbb{Z}} = \{m\tau k\}_{k \leq 0, k \in \mathbb{Z}}$ in Theorems 5 and 6 is equidistant, i.e. it has a simple structure.*

Remark 2 *In Theorem 6, \tilde{f} can be viewed as a result of an arbitrarily small adjustment of f . This \tilde{f} is uniquely defined by sample with a sampling distance $m\tau$, where τ is a distance smaller than the distance defined by critical Nyquist rate for f . Since the value ε can be arbitrarily small, and m and N can be arbitrarily large in Theorems 6 and 4, one can say that the restriction on the sampling rate defined by the Nyquist rate is bypassed, in a certain sense.*

The case of non-bandlimited continuous time functions

Technically speaking, the classical sampling theorem is applicable to band-limited continuous time functions only. However, its important role in signal processing is based on applicability to more general functions since they allow band-limited approximations: any $f \in L_1(\mathbf{R}) \cap L_2(\mathbf{R})$ can be approximated by bandlimited functions $f_\Omega \triangleq \mathcal{F}^{-1}(F\mathbb{I}_{[-\Omega, \Omega]})$ with $\Omega \rightarrow +\infty$, where $F = \mathcal{F}f$, and where \mathbb{I} is the indicator function. However, the sampling frequency has to be increased along with Ω : for a given $\tau > 0$, the sample $\{f_\Omega(\tau k)\}_{k \in \mathbb{Z}}$ defines the function f_Ω if $\tau \leq \pi/\Omega$. Therefore, there is a problem of representation of general functions via sparse samples. A related problem is aliasing of continuous time processes after time discretization. Theorem 6 allows to overcome this obstacle in a certain sense as the following.

Corollary 1 *For any $f \in L_1(\mathbf{R}) \cap L_2(\mathbf{R})$, $\varepsilon > 0$, and $\Delta > 0$, there exist $\bar{\Omega} > 0$ such that the following holds for any $\Omega \geq \bar{\Omega}$:*

$$(i) \sup_{t \in \mathbf{R}} |f(t) - f_\Omega(t)| \leq \varepsilon, \text{ where } f_\Omega = \mathcal{F}^{-1}F_\Omega, F_\Omega = F\mathbb{I}_{[-\Omega, \Omega]}, F = \mathcal{F}f.$$

(ii) The function f_Ω belongs to $L_2^{BL,\Omega}(\mathbf{R})$, and, for $\tau = \pi/\Omega$, $t_k = \tau k$, satisfies assumptions of Theorem 6. For this function, for any $\varepsilon > 0$ there exists $\tilde{f} \in L_2^{BL,\Omega}(\mathbf{R})$ such that $\sup_{t \in \mathbf{R}} |f_\Omega(t) - \tilde{f}(t)| \leq \varepsilon$ and that \tilde{f} is uniquely defined by the values $\{\tilde{f}(\theta_k)\}_{k \in \mathbb{Z}}$ for an equidistant sequence of sampling points $\theta_k = k\Delta$, $k \in \mathbb{Z}$, for any $s \in \mathbb{Z}$.

3 Proofs

To proceed with the proof of the theorems, we will need to prove some preliminary lemma first.

Lemma 1 *Let $x \in V(\delta, \nu m)$ for some $\delta > 0$ and for $m, \nu \in \mathbb{Z}$ such that $m, \nu \geq 1$. Then the following holds for $\varkappa = 1$ or $\varkappa = m$:*

- (i) *For any $n \geq 0$, $n \in \mathbb{Z}$, the sequence $\{x(\varkappa k + n)\}_{k \in \mathbb{Z}}$ is uniquely defined by the values $\{x(\varkappa k + n)\}_{k \in \mathbb{Z}^-}$.*
- (ii) *For any $n \leq 0$, $n \in \mathbb{Z}$, the sequence $\{x(\varkappa k - n)\}_{k \in \mathbb{Z}}$ is uniquely defined by the values $\{x(\varkappa k + n)\}_{k \in \mathbb{Z}^+}$.*

Proof of Lemma 1. It suffices to proof the theorem for the case of the extrapolation from the set \mathbb{Z}^- only; the extension on the extrapolation from \mathbb{Z}^+ is straightforward.

Consider a transfer functions and its inverse Z-transform

$$\hat{H}(z) \triangleq z^n V(z^{\nu m})^n, \quad \hat{h} = \mathcal{Z}^{-1} \hat{H}, \quad (5)$$

where

$$V(z) \triangleq 1 - \exp \left[-\frac{\gamma}{z + 1 - \gamma^{-r}} \right],$$

and where $r > 0$ and $\gamma > 0$ are parameters. This function V was introduced in [3]. (In the notations from [3], $r = 2\mu/(q - 1)$, where $\mu > 1$, $q > 1$ are the parameters). We assume that r is fixed and consider variable $\gamma \rightarrow +\infty$.

In the proof below, we will show that $\hat{H}(e^{i\omega})$ approximates $e^{in\omega}$ and therefore defines a linear n -step predictor with the kernel \hat{h} .

Let $\alpha = \alpha(\gamma) \triangleq 1 - \gamma^{-r}$. Clearly, $\alpha = \alpha(\gamma) \rightarrow 1$ as $\gamma \rightarrow +\infty$.

Let $W(\alpha) = \arccos(-\alpha)$, let $D_+(\alpha) = (-W(\alpha), W(\alpha))$, and let $D(\alpha) \triangleq [-\pi, \pi] \setminus D_+(\alpha)$. We have that $\cos(W(\alpha)) + \alpha = 0$, $\cos(\omega) + \alpha > 0$ for $\omega \in D_+(\alpha)$, and $\cos(\omega) + \alpha < 0$ for $\omega \in D(\alpha)$.

It was shown in [3] that the following holds:

- (i) $V(z) \in H^\infty(D^c)$ and $zV(z) \in H^\infty(D^c)$.

(ii) $V(e^{i\omega}) \rightarrow 1$ for all $\omega \in (-\pi, \pi)$ as $\gamma \rightarrow +\infty$.

(iii) If $\omega \in D_+(\alpha)$ then $\operatorname{Re}\left(\frac{\gamma}{e^{i\omega} + \alpha}\right) > 0$ and $|V(e^{i\omega}) - 1| \leq 1$.

The definitions imply that there exists $\gamma_0 > 0$ such that

$$\sup_{\omega \in (-\pi, \pi] \setminus J_\delta} |V(e^{i\omega}) - 1| \leq 1, \quad \forall \gamma : \gamma > \gamma_0. \quad (6)$$

Without a loss of generality, we assume below that $\gamma > \gamma_0$.

Let

$$Q(\alpha) = \cup_{k=0}^{\nu m-1} \left(\frac{W(\alpha) + 2\pi k}{\nu m}, \frac{2\pi - W(\alpha) + 2\pi k}{\nu m} \right), \quad Q_+(\alpha) = [-\pi, \pi] \setminus Q(\alpha).$$

From the properties of V , it follows that the following holds.

(i) $V(z^{\nu m}) \in H^\infty(D^c)$ and $zV(z^{\nu m}) \in H^\infty(D^c)$.

(ii) $V(e^{i\omega\nu m}) \rightarrow 1$ and $\widehat{H}(e^{i\omega}) \rightarrow e^{in\omega}$ for all $\omega \in (-\pi, \pi] \setminus \{s_{k,\nu m}\}_{k=0}^{\nu m-1}$ as $\gamma \rightarrow +\infty$.

(iii)

$$\operatorname{Re}\left(\frac{\gamma}{e^{i\omega\nu m} + \alpha}\right) > 0, \quad |V(e^{i\omega\nu m}) - 1| \leq 1, \quad \omega \in Q_+(\alpha). \quad (7)$$

Figure 1 shows an example of the shape of error curves for approximation of the forward one step shift operator. More precisely, they show the shape of $|\widehat{H}(e^{i\omega}) - e^{in\omega}|$ for the transfer function (5) with $n = \nu = 1$, and the shape of the corresponding predicting kernel \widehat{h} with some selected parameters.

By the choice of V , it follows that $|V(z)| \rightarrow 0$ as $|z| \rightarrow +\infty$. Hence $v(0) = 0$ for $v = \mathcal{Z}^{-1}V$ and

$$V(z^{\nu m}) = z^{-\nu m}v(1) + z^{-2\nu m}v(2) + z^{-3\nu m}v(3) \dots$$

Clearly, we have that

$$V(z^{\nu m})^n = z^{-n\nu m}w(1) + z^{-(n+1)\nu m}w(2) + z^{-(n+2)\nu m}w(3) \dots,$$

where $w = \mathcal{Z}^{-1}(V(z^{\nu m})^n)$. Hence

$$\begin{aligned} \widehat{H}(z^{\nu m}) &= z^n V(z^{\nu m})^n = z^{n-n\nu m}w(1) + z^{n-(n+1)\nu m}w(2) + z^{n-(n+2)\nu m}w(3) \dots \\ &= z^{n-n\nu m}\widehat{h}(n\nu m - n) + z^{n-(n+1)\nu m}\widehat{h}((n+1)\nu m - n) + z^{n-(n+2)\nu m}\widehat{h}((n+2)\nu m - n) \dots \end{aligned}$$

In particular, $\widehat{H} \in H^\infty(D^c)$ and

$$\widehat{h}(k) = 0 \quad \text{if either} \quad (k+n)/m \notin \mathbb{Z} \quad \text{or} \quad k < mn - n. \quad (8)$$

(In fact, the sequence \hat{h} is more sparse for $n > 1$ or $\nu > 1$ than is shown in (8); nevertheless, (8) is sufficient for our purposes).

It can be noted that \hat{h} is real valued, since $\widehat{H}(\bar{z}) = \overline{\widehat{H}(z)}$.

Assume that we are given $x \in V(\delta, \nu m)$. Let $X \triangleq \mathcal{Z}x$, $Y(z) = zX(z)$, and $\hat{Y} \triangleq \mathcal{Z}\hat{y} = \widehat{H}X$.

By the notations accepted above, it follows that B_1 is the unit ball in ℓ_2 . Let us show that

$$\sup_{k \in \mathbb{Z}} |x(k+n) - \hat{y}(k)| \rightarrow 0 \quad \text{as } \gamma \rightarrow +\infty \quad \text{uniformly over } x \in V(\delta, \nu m) \cap B_1, \quad (9)$$

where the process \hat{y} is the output of a linear predictor defined by the kernels \hat{h} as

$$\hat{y}(k) \triangleq \sum_{p=-\infty}^k \hat{h}(k-p)x(p). \quad (10)$$

Remark 3 By (8),(10), an estimate $\hat{y}(k)$ of $x(k+n)$ is constructed using the observations $\{x(k+n-qm)\}_{q \geq n, q \in \mathbb{Z}}$. This can be seen from representation of convolution (10) as

$$y(k) = \sum_{j \in \mathbb{Z}, j \geq n} \hat{h}(jm-n)x(k-jm+n).$$

This predictor produces the process $\hat{y}(k)$ approximating $x(k+n)$ as $\gamma \rightarrow +\infty$ for all inputs $x \in V(\delta, \nu m)$.

Let $y(t) \triangleq x(t+n)$ and $Y = \mathcal{Z}y$. We have that $\|Y(e^{i\omega}) - \hat{Y}(e^{i\omega})\|_{L_1(-\pi, \pi)} = I_1 + I_2$, where

$$I_1 = \int_{Q(\alpha)} |Y(e^{i\omega}) - \hat{Y}(e^{i\omega})| d\omega, \quad I_2 = \int_{Q_+(\alpha)} |Y(e^{i\omega}) - \hat{Y}(e^{i\omega})| d\omega.$$

By the definitions, there exists $\gamma_0 > 0$ such that, for all $\gamma > \gamma_0$, we have that

$$Q(\alpha) \subset J_{\delta, m}, \quad I_1 = 0. \quad (11)$$

This means that $I_1 \rightarrow 0$ as $\gamma \rightarrow +\infty$ uniformly over $x \in V(\delta, \nu m) \cap B_1$.

Let us estimate I_2 . We have that

$$\begin{aligned} I_2 &= \int_{Q_+(\alpha)} |(e^{i\omega n} - H(e^{i\omega}))X(e^{i\omega})| d\omega \leq \int_{Q_+(\alpha)} |e^{i\omega n}(1 - V(e^{i\omega\nu m})^n)X(e^{i\omega})| d\omega \\ &\leq \|1 - V(e^{i\omega\nu m})^n\|_{L_2(Q_+(\alpha))} \|X(e^{i\omega})\|_{L_2(-\pi, \pi)}. \end{aligned}$$

Further, $\mathbb{I}_{Q_+(\alpha)}(\omega)|1 - V(e^{i\omega\nu m})^n| \rightarrow 0$ a.e. as $\gamma \rightarrow +\infty$. By (6),

$$\mathbb{I}_{Q_+(\alpha)}(\omega)|1 - V(e^{i\omega\nu m})^n| \leq 2^n, \quad \forall \gamma : \gamma > \gamma_0.$$

From Lebesgue Dominance Theorem, it follows that $\|1 - V(e^{i\omega\nu m})^n\|_{L_2(Q_+(\alpha))} \rightarrow 0$ as $\gamma \rightarrow +\infty$. It follows that $I_1 + I_2 \rightarrow 0$ uniformly over $x \in V(\delta, \nu m) \cap B_1$. Hence (9) holds and

$$\sup_{k \in \mathbb{Z}} |x(k+n) - \hat{y}(k)| \rightarrow 0 \quad \text{as } \gamma \rightarrow +\infty \quad \text{uniformly over } x \in V(\delta, \nu m) \cap B_1.$$

Hence the predicting kernels $\hat{h}(\cdot) = \mathcal{Z}^{-1}\hat{H}$ are such as required.

To complete the proof of Lemma 1, it suffices to prove that if x is such that $x(k\kappa + n) = 0$ for $k \leq s$, then $x(k\kappa + n) = 0$ for $k > q$ for all $q \in \mathbb{Z}$. By the predictability established by (9) for a process $x \in V(\delta, \nu m)$ such that $x(k\kappa + n) = 0$ for $k \leq s$, it follows that $x(k\kappa + \kappa + n) = 0$; we obtain this by letting $\gamma \rightarrow +\infty$ in (5). (See also Remark 3). Similarly, we obtain that $x(k\kappa + 2\kappa + q) = 0$. Repeating this procedure n , we obtain that $x(k\kappa + n) = 0$ for all $k \in \mathbb{Z}$. This completes the proof of Lemma 1. \square

Remark 4 Assume that $\{\hat{x}_d\}_{d=-m+1}^{m-1} \in \mathcal{U}(\delta, m)$ for some $\delta > 0$.

- (i) Lemma 1 applied with $\kappa = 1$ implies that all \hat{x}_d are uniquely defined by $\{\hat{x}_0(k)\}_{k \in \mathbb{Z}^-}$ for $d = 0, 1, \dots, m-1$.
- (ii) Lemma 1 applied with $\kappa = m$ and $n = 0$ implies that all values $\hat{x}_d(mk)$ for $k > 0$ are uniquely defined by the sequence $\{\hat{x}_0(mk)\}_{k \in \mathbb{Z}^+}$ for $d = -m+1, 1, \dots, 1$.

Proof of Theorem 1. Clearly, $\tilde{x} \in \ell_2$, since $\|\tilde{x}\|_{\ell_2} \leq \sum_{d=-m+1}^{m-1} \|\hat{x}_d\|_{\ell_2}$. Let $d \in \{0, 1, \dots, m-1\}$. By Lemma 1 applied with $\kappa = m$, we have that a sequence $\{\hat{x}_d(mk)\}_{k \geq 1}$ is uniquely defined by the sequence $\{\hat{x}_d(mk)\}_{k \leq 0}$ (see Remark 3 above). Since $\{\tilde{x}(mk-d)\}_{k \geq 1} = \{\hat{x}_d(mk)\}_{k \geq 1}$ and $\{\hat{x}_d(mk)\}_{k \leq 0} = \{\tilde{x}(mk)\}_{k \leq 0}$, this implies that $\{\tilde{x}(mk-d)\}_{k \geq 1}$ is uniquely defined by the sequence $\{\tilde{x}(mk)\}_{k \leq 0}$.

Similar reasoning can be applied to the case where $d \in \{-m+1, \dots, 1\}$. By Lemma 1 applied with $\kappa = m$, we have that a sequence $\{\hat{x}_d(mk)\}_{k \leq -1}$ is uniquely defined by the sequence $\{\hat{x}_d(mk)\}_{k \geq 0}$. Since $\{\tilde{x}(mk-d)\}_{k \leq -1} = \{\hat{x}_d(mk)\}_{k \geq 0}$ and $\{\hat{x}_d(mk)\}_{k \geq 0} = \{\tilde{x}(mk)\}_{k \geq 0}$, this implies that $\{\tilde{x}(mk-d)\}_{k \leq -1}$ is uniquely defined again by the sequence $\{\tilde{x}(mk)\}_{k \geq 0}$. This completes the proof of Theorem 1. \square

Proof of Theorem 2. For $m, d \in \mathbb{Z}$, let $\mu(m, d) \triangleq 2^d m$ for $d \geq 0$, and let $\mu(m, d) = 2^{2m+d-1} m$ for $d < 0$.

Let $s_{\mu(m,d),k}$ and $J_{\delta,\mu(m,d)}$ be defined by (1). Let us show that, for small enough $\delta > 0$, the sets $J_{\delta,\mu(m,d)}$ are disjoint for different $d \in \{-m-1, \dots, m-1\}$. It suffices to show that $s_{\mu(m,d_1),k} \neq s_{\mu(m,d_2),l}$ if $d_1 \neq d_2$ for all $k, l \in \mathbb{Z}$. Suppose that $s_{\mu(m,d_1),k} = s_{\mu(m,d_2),l}$ for some $k, l \in \mathbb{Z}$ for $d_2 = d_1 + r$, for some $d_1, d_2 \in \{-m+1, \dots, m-1\}$ such that $r \triangleq d_2 - d_1 > 0$. In this case, the definitions imply that

$$\frac{2k-1}{2^{d_1}} = \frac{2l-1}{2^{d_2}}.$$

This means that $2^r(2k-1) = 2l-1$, and, therefore the number $2^r(2k-1)$ is odd. This is impossible since we had assumed that $r > 0$. Hence the sets $J_{\delta,\mu(m,d)}$ are disjoint for different d for small $\delta > 0$.

Without a loss of generality, we assume that all δ used below are small enough to ensure that this property hold.

Let $Y_d \triangleq X_0 - X_d$ and $y_d = \mathcal{Z}^{-1}Y_d$, where $X_d = \mathcal{Z}x_d$, $d = -m+1, \dots, m-1$. Since $x_d(k) = x_0(k)$ for $k \leq 0$ and $d > 0$, and $x_d(k) = x_0(k)$ for $k \geq 0$ and $d < 0$, it follows that $y_d(k) = 0$ for $k \leq 0$ and $d > 0$, and $y_d(k) = 0$ for $k \geq 0$ and $d < 0$. $d = 1, 2, \dots, m-1$. In addition, $Y_0 \equiv 0$ and $y_0 \equiv 0$.

Further, let $\mathcal{T} = \{k \in \mathbb{Z} : |k| \leq m-1, k \neq 0\}$, and let

$$\widehat{X}_0(e^{i\omega}) \triangleq X_0(e^{i\omega}) \mathbb{I}_{\{\omega \notin \bigcup_{d=-m+1}^{m-1} J_{\delta, \mu(m, d)}\}} + \sum_{d \in \mathcal{T}} Y_d(e^{i\omega}) \mathbb{I}_{\{\omega \in J_{\delta, \mu(m, d)}\}},$$

and let $\widehat{X}_d \triangleq \widehat{X}_0 - Y_d$ for $d \in \mathcal{T}$. By the definitions,

$$\begin{aligned} \widehat{X}_d(e^{i\omega}) &= \widehat{X}_0(e^{i\omega}) - X_0(e^{i\omega}) + X_d(e^{i\omega}) \\ &= X_d(e^{i\omega}) + \sum_{p \in \mathcal{T}} (Y_p(e^{i\omega}) - X_0(e^{i\omega})) \mathbb{I}_{\{\omega \in J_{\delta, \mu(m, p)}\}} \\ &= X_d(e^{i\omega}) - \sum_{p \in \mathcal{T}} X_p(e^{i\omega}) \mathbb{I}_{\{\omega \in J_{\delta, \mu(m, p)}\}} \\ &= X_d(e^{i\omega}) \mathbb{I}_{\{\omega \notin J_{\delta, \mu(m, d)}\}} - \sum_{p \in \mathcal{T}, p \neq d} X_p(e^{i\omega}) \mathbb{I}_{\{\omega \in J_{\delta, \mu(m, p)}\}}, \quad d = -m+1, \dots, m-1. \end{aligned}$$

Since the sets $J_{\delta, \mu(m, d)}$ are mutually disjoint, it follows that $\widehat{X}_d(e^{i\omega}) = 0$ for $\omega \in J_{\delta, \mu(m, d)}$ and for all d . Hence $\widehat{x}_d \triangleq \mathcal{Z}^{-1}\widehat{X}_d \in V(\delta, \mu(m, d))$ for all d . It follows that the branching process $\{\widehat{x}_d\}_{d=-m+1}^{m-1}$ belongs to $\cup_{\delta>0} \mathcal{U}(\delta, m)$. In particular, $\widehat{x}_d(k) = \widehat{x}_0(k)$, for $k \leq 0$ and $d > 0$, and $\widehat{x}_d(k) = \widehat{x}_0(k)$, for $k \geq 0$ and $d < 0$. Clearly, $\|\widehat{x}_d - x_d\|_{\ell_2} \rightarrow 0$ as $\delta \rightarrow 0$ for $d = -m+1, \dots, m-1$. This completes the proof of Theorem 2. \square

Proof of Theorem 3. Let $\{x_d\}_{d=-m+1}^{m-1}, \{\widehat{x}_d\}_{d=-m+1}^{m-1} \in \cup_{\delta>0} \mathcal{U}(\delta, m)$ be such that suggested in the theorem's statement. By the definitions, it follows that

$$\widetilde{x}(k) - x(k) = \widehat{x}_d(k+d) - x_d(k+d), \quad k \geq 0, \quad d \geq 0, \quad (k+d)/m \in \mathbb{Z},$$

and

$$\widetilde{x}(k) - x(k) = \widehat{x}_d(k+d) - x_d(k+d), \quad k < 0, \quad d < 0, \quad (k+d)/m \in \mathbb{Z}.$$

Then (4) follows from (2). This completes the proof of Theorem 3. \square

Proof of Theorem 4. It suffices to show that the error for recovery a single term $\widetilde{x}(n)$ for a given integer n from the sequence $\{\widetilde{x}(mk)\}_{k \in \mathbb{Z}, k \leq 0}$ can be made arbitrarily small is a well-posed problem; the proof for a finite set of values to recover is similar. Furthermore, it suffices to consider $n > 0$; the case of $n < 0$ can be considered similarly.

Let us consider an input sequence $x \in \ell_2$ such that

$$x = \tilde{x} + \eta, \quad (12)$$

where $\eta \in \ell_2$ represents a noise, and where $\tilde{x} \in U(\delta, m)$ is the representative branch for a branching process $\{\hat{x}_d\}_{d=-m+1}^{m-1} \in \mathcal{U}(\delta, m)$.

Let $\eta_n \in \ell_2$ be defined such that $\eta_n(k) = \eta(k)\mathbb{I}_{\{k \leq n\}}$. Let $N \triangleq \mathcal{Z}\eta_n$, and let $\sigma = \|N(e^{i\omega})\|_{L_2(-\pi, \pi)}$; this parameter represents the intensity of the noise. Let $X = \mathcal{Z}x$, $\hat{X}_d = \mathcal{Z}\hat{x}_d$, and $N = \mathcal{Z}\eta$.

Let $d \in \{0, 1, \dots, m-1\}$ be such that $(n+d)/m \in \mathbb{Z}$.

Let $\delta > 0$ be given and an arbitrarily small $\varepsilon > 0$ be given. Assume that the parameters (γ, r) of V and \hat{H} in (5) are selected such that

$$\sup_{k \in \mathbb{Z}} |\bar{x}(k+n) - \bar{y}(k)| \leq \varepsilon/2 \quad \forall \bar{x} \in V(\delta, \mu(m, d)) \cap B_1, \quad (13)$$

for $\bar{y} = \mathcal{Z}^{-1}(\hat{H}\bar{X})$ and $\bar{X} = \mathcal{Z}\bar{x}$. Here \hat{H} is selected by (5), and B_1 is the unit ball in ℓ_2 .

By the choice of d , we have that there exists $p \in \mathbb{Z}$, $p \geq 1$, such that $n = pm - d$ and

$$\hat{x}_d(pm) = \tilde{x}(pm - d) = \tilde{x}(n). \quad (14)$$

By (8) and Remark 3, the kernel \hat{h} produces an estimate y_d of $\hat{x}_d(pm)$ based on observations of $\{\hat{x}_d(km)\}_{k < 0}$.

Let us assume first that $\sigma = 0$. In this case, we have that

$$\hat{E} \triangleq |\hat{y}_d(0) - \hat{x}_d(pm)| = |\hat{y}_d(0) - \tilde{x}(n)| \leq \varepsilon/2. \quad (15)$$

Let us consider the case where $\sigma > 0$. In this case, we have that

$$|\hat{y}_d(0) - \hat{x}_d(pm)| = |\hat{y}_d(0) - x(n)| \leq \hat{E} + E_\eta,$$

where

$$E_\eta \triangleq \frac{1}{2\pi} \|(\hat{H}(e^{i\omega}) - e^{i\omega n})N(e^{i\omega})\|_{L_1(-\pi, \pi)} \leq \sigma(\kappa + 1),$$

and where

$$\kappa \triangleq \sup_{\omega \in [-\pi, \pi]} |\hat{H}(e^{i\omega})|.$$

Assume that $\eta(k) = \xi(k)\mathbb{I}_{\{-N \leq k \leq n\}} - \hat{x}_0(k)\mathbb{I}_{\{k < -N\}}$, for an integer $N > 0$. In this case, (12) gives that $x(k) = (x_0(k) + \xi(k))\mathbb{I}_{\{k > -N\}}$ for $k \leq n$. In addition, we have in this case that $\sigma \rightarrow 0$ as $N \rightarrow +\infty$ and $\|\xi\|_{\ell_2(-N, 0)} \rightarrow 0$. If N is large enough and $\|\xi\|_{\ell_2(-N, 0)}$ is small enough such that $\sigma(\kappa + 1) < \varepsilon/2$, then $\|\hat{y}_d(0) - x(n)\|_{\ell_\infty} \leq \varepsilon$. This completes the proof of Theorem 4. \square

Remark 5 By the properties of \widehat{H} , we have that $\kappa \rightarrow +\infty$ as $\gamma \rightarrow +\infty$. This implies that error (15) will be increasing if $\widehat{\varepsilon} \rightarrow 0$ for any given $\sigma > 0$. This means that, in practice, the predictor should not target too small a size of the error, since it is impossible to ensure that $\sigma = 0$ due to inevitable data truncation.

Proof of Theorem 5. The previous proof shows that $\tilde{x} \in \ell_2$. For $\tilde{X} = \mathcal{Z}\tilde{x}$, we have that

$$\tilde{x}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \tilde{X}(e^{i\nu}) e^{i\nu k} d\nu = \frac{\tau}{2\pi} \int_{-\pi/\tau}^{\pi/\tau} \tilde{X}(e^{i\tau\omega}) e^{i\omega\tau k} d\omega,$$

with the change of variables $\nu = \tau\omega$. Let us define function $\tilde{F} : i\mathbf{R} \rightarrow \mathbf{C}$ as $\tilde{F}(i\omega) \triangleq \tau\tilde{X}(e^{i\tau\omega})$ for $\omega \in \mathbf{R}$. Then

$$\tilde{x}(k) = \frac{1}{2\pi} \int_{-\pi/\tau}^{\pi/\tau} \tilde{F}(i\omega) e^{i\omega\tau k} d\omega.$$

Since $\tilde{X}(e^{i\cdot}) \in L_2(-\pi, \pi)$, this implies that $\tilde{F}(i\cdot) \in L_2(i\mathbf{R})$. The sequence $\{\tilde{x}(k)\}_{k \in \mathbb{Z}}$ represents the sequence of Fourier coefficients of \tilde{F} and defines \tilde{F} uniquely. By Theorem 1, this sequence is uniquely defined by the sequence $\{\tilde{x}(mk)\}_{k \in \mathbb{Z}}$. Let $\tilde{f} \triangleq \mathcal{F}\tilde{F}$. Clearly, $\tilde{f} \in L_2^{BL, \pi/\tau}(\mathbf{R})$ and it is uniquely defined by the sequence $\{\tilde{x}(mk)\}_{k \in \mathbb{Z}}$. This completes the proof of Theorem 5. \square

Proof of Theorem 6. We use notations from the proofs above. The existences of the required $\{\widehat{x}_d\}_{d=-m+1}^{m-1}$, \tilde{x} , and \tilde{f} , follows from Theorem 1-2. We have for $F = \mathcal{F}f$ that, for some $\bar{C} > 0$, that

$$\|\tilde{F}(i\cdot) - F(i\cdot)\|_{L_2(\mathbf{R})} = \sqrt{\tau} \|\tilde{x} - x\|_{\ell_2} \leq \sqrt{\tau} \sum_{d=-m+1}^{m-1} \|\widehat{x}_d - x_d\|_{\ell_2} \leq \sqrt{\tau}(2m-1)\varepsilon.$$

Since $F(i\omega) = 0$ and $\tilde{F}(i\omega) = 0$ if $|\omega| > \pi/\tau$, it follows that

$$\|\tilde{F}(i\cdot) - F(i\cdot)\|_{L_1(\mathbf{R})} \leq \sqrt{2\Omega} \|\tilde{F}(i\cdot) - F(i\cdot)\|_{L_2(\mathbf{R})}.$$

Combining these estimates, we obtain that

$$\|\tilde{f} - f\|_{L_p(\mathbf{R})} \leq 2\pi \max(1, \sqrt{2\Omega}) \|\tilde{F}(i\cdot) - F(i\cdot)\|_{L_2(\mathbf{R})} \leq 2\pi \max(1, \sqrt{2\Omega}) \sqrt{\tau}(2m-1)\varepsilon.$$

This completes the proof of Theorem 6. \square

Proof of Theorem 7. It suffices to observe that $\tilde{x}(k) = \tilde{f}(t_k)$ satisfy the assumptions of Theorem 7. \square

Proof of Corollary 1. We have that

$$\begin{aligned} \sup_{t \in \mathbf{R}} |f(t) - f_{\Omega}(t)| &\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |F(i\omega) - F_{\Omega}(i\omega)| d\omega \\ &= \frac{1}{2\pi} \int_{\mathbf{R} \setminus [-\Omega, \Omega]} |F(i\omega)| d\omega \rightarrow 0 \quad \text{as } \Omega \rightarrow +\infty, m \rightarrow +\infty. \end{aligned}$$

Then statement (i) follows. The function f_{Ω} has the same properties as the function f in Theorem 6.

This completes the proof of Corollary 1. \square

4 On numerical implementation

Theorems 6 and 4 allow to bypass, in a certain sense, the restriction on the sampling rate defined by the critical Nyquist rate. Admittedly, this is only the very first step in solution of the problem. The proofs of the results implies a recovering algorithm. However, numerical implementation to data compression and recovery using the scheme in the proof above is numerically challenging. So far, this possibility is rather theoretical, since the presence of a noise in the data or in the measurements can cause a significant error if $\gamma \rightarrow +\infty$. The aims of the present paper are limited by the theoretical aspects of possibility of recovering functions from decimated samples; we leave analysis of possibilities for numerical implementation for the future research. However, let us summarize briefly some steps require for the numerical implementation.

An algorithm for compression and recovery

Assume that we are given $f \in L_1(\mathbf{R}) \cap L_2(\mathbf{R})$. Theorems 6-7 and Corollary 1 offer a method of compressed representation of f via a sparse sample of some function \tilde{f} that is close enough to f ; this can be considered as a compression problem.

As is described in Corollary 1, f can be approximated by bandlimited functions $f_\Omega \triangleq \mathcal{F}^{-1}(F(i\omega)\mathbb{I}_{[-\Omega, \Omega]}(\omega))$, $F = \mathcal{F}f$, where $\Omega \rightarrow +\infty$. The classical Sampling Theorem allows to restore the Fourier transform f_Ω from the sampling series $\{f_\Omega(t_k)\}_{k \in \mathbb{Z}}$, where $t_k = \tau k$, $k \in \mathbb{Z}$, $\tau \in (0, \pi/\Omega)$. The aim is a compressed representations using the sampling times mt_k , where $m > 0$ defines a desired sparsity. For this, we suggest to approximate f_Ω by a function \tilde{f} such as described in Theorem 6-4. The corresponding steps for this compression algorithm can be summarized as follows.

- (C1) Select $\Omega > 0$ to ensure sufficient approximation of f by f_Ω described in Corollary 1. Select $\tau < \pi/\Omega$ and define $x_0(k) = f(t_k)$, where $t_k = k\tau$, $k \in \mathbb{Z}$. Select a integer $m > 0$ such that $m\tau k$ is a distance that ensures sufficient sparsity.
- (C2) Define $x_d = \mathcal{M}_d x_0$, where $d = -m + 1, \dots, m - 1$.
- (C3) Define \hat{X}_d as described in the proof of Theorem 2 using sufficiently small $\delta > 0$. Find $\hat{x}_d = \mathcal{Z}^{-1} \hat{X}_d$; these sequences form a branching process $\{\hat{x}_d\}_{d=-m+1}^{m-1}$. Find the representative branch \tilde{x} by Definition 2.

Any of steps (C1)-(C3) represents a well-posed problem.

The sequence $\{\tilde{x}(mk)\}_{k \in \mathbb{Z}}$ can be accepted as a compressed representation of f . This can be justified as the following. The sequence \tilde{x} is uniquely defined by the sequence $\{\tilde{x}(mk)\}_{k \in \mathbb{Z}}$, and there

exists a unique $\tilde{f} \in L_2^{BL, \pi/\tau}(\mathbf{R})$ such that $\tilde{f}(t_k) = \tilde{x}(k)$ for all $k \in \mathbb{Z}$. This \tilde{f} approximates f_Ω given that δ is sufficiently small. In addition, \tilde{f} is uniquely defined by \tilde{x} and hence by $\{\tilde{x}(mk)\}_{k \in \mathbb{Z}}$.

Recovery \tilde{f} from the suggested compressed representation $\{\tilde{x}(mk)\}_{k \in \mathbb{Z}}$ would require the following steps.

(R1) Recover \tilde{x} from the sequence $\{\tilde{x}(mk)\}_{k \in \mathbb{Z}}$.

(R2) Recover \tilde{f} from \tilde{x} .

Step (R2) represents a well-posed problem which solution is implied by the Sampling Theorem. In addition, Theorem 4 establishes some well-posedness of Step (R1) for recovery of a finite set of values.

Some examples of numerical experiments for the predicting algorithm assumed by Step (R1) can be found in [7] (for the case where $m = n = 1$, in the notations of the present papers).

5 Discussion and future developments

The paper suggests a uniqueness result for sparse sampling (Theorem 6 (ii)) and a version of the Sampling Theorem where the restriction on the sampling rate defined by the Nyquist rate is bypassed, in a certain sense (Theorem 4). This was only the very first step in attacking the problem; the numerical implementation to data compression and recovery is quite challenging and there are many open questions. In particular, it involves the solution of ill-posed inverse problems. In addition, there are other possible extensions of this work that we will leave for the future research.

- (i) A theoretical problem arises: *How to detect a trace $x_0|_{k \leq 0}$ that can be extended into a branching process featuring degeneracy described in Definition 3 for $d > 0$?* This is actually a non-trivial question even for $m = 1$; see discussion Definition 1 in [6] and discussion in [5, 6, 8].
- (ii) To cover image processing tasks, the approach of this paper has to be extended on functions $f : \mathbf{R}^2 \rightarrow \mathbf{R}$. This would require to extend the predicting algorithms from [3] used in the present papers on processes defined on multidimensional lattices, possibly, using the setting from [18].
- (iii) The result of this paper allows many acceptable modifications such as the following.
 - The choice of mappings \mathcal{M}_d allows many modifications; for example, $x_d = \mathcal{M}_d x$ are defined such that there exists $\theta \geq m$ such that $x_d(k) = x(k - d)$ for $k > \theta$, without any restrictions on $\{x_d(k)\}_{k=1, \dots, \theta}$.
 - Conditions of Theorem 1 can be relaxed: instead of the condition that the spectrum vanishes on open sets $J_{\delta, m}$, we can require that the spectrum vanishes only at the middle points of

the intervals forming $J_{\delta,m}$; however, the rate of vanishing has to be sufficient, similarly to [3]. This is because the predicting algorithm [3] does not require that the spectrum of an underlying process is vanishing on an open subset of \mathbb{T} .

– The choice (5) of predictors presented in the proofs above is not unique. For example, we could use a predicting algorithm from [4] instead of the the algorithm based on [3] used above.

Acknowledgment

This work was supported by ARC grant of Australia DP120100928 to the author.

References

- [1] T. Cai, G. Xu, and J. Zhang. (2009), On recovery of sparse signals via ℓ_1 minimization, *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3388-3397.
- [2] Beurling, A., Malliavin, P. (1967). On the closure of characters and the zeros of entire functions, *Acta Mathematica* **118** 79–93.
- [3] Dokuchaev, N. (2012). Predictors for discrete time processes with energy decay on higher frequencies. *IEEE Transactions on Signal Processing* **60**, No. 11, 6027-6030.
- [4] Dokuchaev, N. (2012). On predictors for band-limited and high-frequency time series. *Signal Processing* **92**, iss. 10, 2571-2575.
- [5] Dokuchaev, N. (2012). Causal band-limited approximation and forecasting for discrete time processes. ArXiv 1208.3278.
- [6] Dokuchaev, N. (2016). On recovering missing values in a pathwise setting. ArXiv: <https://arxiv.org/abs/1604.04967>.
- [7] Dokuchaev, N. (2016). Near-ideal causal smoothing filters for the real sequences. *Signal Processing* **118**, iss. 1, pp. 285-293.
- [8] Dokuchaev, N. (2017). On detecting predictability of one-sided sequences. *Digital Signal Processing* **62** 26-29.
- [9] Ferreira P. G. S. G.. (1992). Incomplete sampling series and the recovery of missing samples from oversampled bandlimited signals. *IEEE Trans. Signal Processing* **40**, iss. 1, 225-227.

- [10] Ferreira P. G. S. G.. (1995). Sampling Series With An Infinite Number Of Unknown Samples. In: *SampTA'95, 1995 Workshop on Sampling Theory and Applications*, 268-271.
- [11] Jerri, A. (1977). The Shannon sampling theorem - its various extensions and applications: A tutorial review. *Proc. IEEE* **65**, 11, 1565–1596.
- [12] Kotelnikov, V.A. (1933). On the carrying capacity of the ether and wire in telecommunications. *Material for the First All-Union Conference on Questions of Communication, Izd. Red. Upr. Svyazi RKKA, Moscow*, 1933.
- [13] Nyquist, H. (1928). Certain topics in telegraph transmission theory, *Trans. AIEE*, Trans., vol. 47, pp. 617-644.
- [14] Landau H.J. (1964). A sparse regular sequence of exponentials closed on large sets. *Bull. Amer. Math. Soc.* **70** 566-569.
- [15] Landau H.J. (1967). Sampling, data transmission, and the Nyquist rate. *Proc. IEEE* **55** (10), 1701-1706.
- [16] Olevski, A., and Ulanovskii, A. (2008). Universal sampling and interpolation of band-limited signals. *Geometric and Functional Analysis*, vol. 18, no. 3, pp. 1029–1052.
- [17] Olevskii A.M. and Ulanovskii A. (2016). *Functions with Disconnected Spectrum: Sampling, Interpolation, Translates*. Amer. Math. Soc., Univ. Lect. Ser. V.46.
- [18] Petersen, D.P. and Middleton D. (1962). Sampling and reconstruction of wave-number-limited functions in N-dimensional Euclidean spaces. *Information and Control*, vol. 5, pp. 279-323.
- [19] Shannon, C.E. (1949). Communications in the presence of noise, *Proc. Institute of Radio Engineers*, vol. 37, no.1, pp. 1021.
- [20] Ulanovskii, A. (2001). On Landaus phenomenon in R^n . *Mathematica Scandinavica*, vol. 88, no. 1, 72–78.
- [21] Ulanovskii, A. (2001). Sparse systems of functions closed on large sets in R^N , *Journal of the London Mathematical Society*, vol. 63, no. 2, 428–440.
- [22] Unser, M. (2000). Sampling 50 years after Shannon, *Proceedings of the IEEE*, vol. 88, pp. 569-587.,
- [23] Vaidyanathan P. P. (1987). On predicting a band-limited signal based on past sample values, in *Proceedings of the IEEE*, vol. 75, no. 8, pp. 1125-1127.

- [24] Vaidyanathan, P.P. (2001). Generalizations of the Sampling Theorem: Seven Decades After Nyquist. *IEEE Transactions on circuits and systemsI: fundamental theory and applications*, v. 48, NO. 9,
- [25] Whittaker, E.T. (1915). On the Functions Which are Represented by the Expansions of the Interpolation Theory. *Proc. Royal Soc. Edinburgh, Sec. A*, vol.35, pp. 181194.
- [26] Zayed A. (1993). *Advances in Shannon's Sampling Theory*. Boca Raton, CRC Press, N.Y, London, Tokyo.

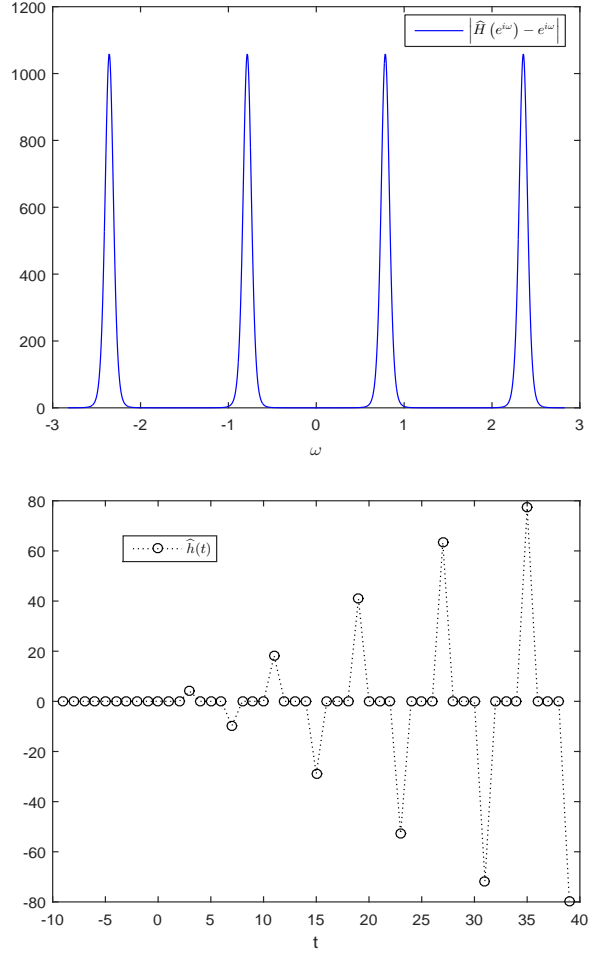


Figure 1: Approximation of the one-step forward shift operator: the values $|\hat{H}(e^{i\omega}) - e^{i\omega}|$ for the transfer function of the predictor (5) and the values of the corresponding kernel $\hat{h} = \mathcal{Z}^{-1}\hat{H}$ with $\gamma = 4$, $r = 0.4$, $m = 4$.